Classification of Retinal Diseases from OCT scans using Convolutional Neural Networks

Suhail Najeeb[1,*], Nowshin Sharmile[1], Md. Sajid Khan[1], Ipsita Sahin[1],
Mohammad Tariqul Islam[2] and Mohammed Imamul Hassan Bhuiyan[1]
[1]Department of Electrical and Electronic Engineering,
Bangladesh Univsersity of Engineering and Technology, Dhaka, Bangladesh
[2]Department of Electrical Engineering, Princeton University
*suhailnajeeb@ieee.org

***Abstract*—Biomedical image classification for diseases is a lengthy and manual process. However recent progresses in computer vision has enabled detection and classification of medical images using machine intelligence a more feasible solution. We explore the possibility of automated detection and classification of retinal abnormalities from retinal OCT scan images of patients. We develop an algorithm to detect the region of interest from a retinal OCT scan and use a computationally inexpensive single layer convolutional neutral network structure for the classification process. Our model is trained on an open source retinal OCT dataset containing 83,484 images of various tunnel disease patients and provides a feasible classification accuracy.**

***Index Terms*—Computer Vision; Retinal OCT Scan; CNN; Biomedical Image Classification.**

I. Introduction

Thousands of people are diagnosed with retinal abnormalities every day. The diagnosis process requires trained medical professionals to check and classify retinal abnormalities from OCT scan images. This is a lengthy and manual process in general. However, alternate solutions could be found to this problem such as developing an automated algorithm that could aid medical professionals beforehand in the detection and classification task.

Deep learning applications in medical imaging is an emerging practice. While trying to classify artery/vein diseases, Girard & Cheriet [1] proposed a novel method for artery/vein classification combining deep learning and graph propagation strategies from the fundus images. Their method obtained an accuracy of 93.3% on the DRIVE database compared to the state of the art accuracy of 91.7%. Gleryz & Ulusoy [2] suggested that retinal vessel segmentation and extracting features such as tortuosity, width, length related to those vessels can be used in diagnosis, treatment and screening of many diseases such as retinopathy of prematurity, hypertension and diabetes. They proposed a convolutional neural networks (CNN) for automatic segmentation of retinal vessels. The proposed CNN model is tested on DRIVE dataset and achieved a better performance than literature.

Sambaturu et al. [3] utilized a region-based CNN approach to automatically mark the exudates and hemorrhages in a fundus image for studying diabetic retinopathy. Their approach extracted the characteristic region proposals characterizing the disease and successfully marked the lesions with a recall of 90%. Calimeri, Marzullo, Stamile & Terracina [4] presented a supervised method for automatically detecting the position of the Optic Disc in retinal fundus digital images and the goal has been achieved by means of a proper reuse of previous knowledge from a pre-trained Convolutional Neural Network (CNN), already able to detect faces in an image. Experimental analyses showed high level of accuracy in the detection of the optic disc on the DRIVE, STARE and DRIONS databases.

Triwijoyo et al. [5] experimented on retinal images to assess eye diseases such as glaucoma and hypertension. They suggested that understanding vascular abnormalities from retinal image helps medical doctors in providing early diagnosis and treatment to stroke, cerebral damage, carotid atherosclerosis, artery disease, cerebral amyloid angiopathy etc. They took Convolutional Neural Networks (CNN) as a classifier to recognize retinal images. They used the public STARE color image dataset which is categorized into 15 classes. The experimentation showed that the CNN model can achieve an accuracy of 80.93%.

Athar, Vahadane, Joshi & Dastidar [6] reviewed Retinal Optical Coherence Tomography (OCT) scans to detect a common type of abnormality found in retina which is a Fluid Filled Region (FFR). They presented a method to simultaneously classify and localize FFRs within retinal OCT scans using a specialized Convolutional Neural Network (CNN). They compared different architectures to see which ones give us the best localization and classification metrics. They found that architectures using Dense Blocks and Scaled Exponential Unit (SeLU) activations gave the best localizations with a Mean Average Precision (mAP) of 0.75 on true positive images and a classification accuracy of 94.8%. Gopinath & Sivaswamy [7] were interested in the early detection of retinal diseases from the automated and accurate segmentation of cystoid structures in Optical Coherence Tomography (OCT). They proposed a novel method for localizing cysts in 3D OCT volumes which is biologically inspired and based on selective enhancement of the cysts, by inducing motion to a given OCT slice. A Convolutional Neural Network (CNN) was designed by them to learn a mapping function that combined the result of multiple such motions to produce a probability map for cyst locations in a given slice. The final segmentation of cysts was obtained via simple clustering of the detected cyst locations. Their proposed system outperformed all previous benchmarks. Review of the literature reflects the strength of computer vision and machine intelligence in biomedical image processing tasks.

978-1-5386-5541-2/18/$31.00

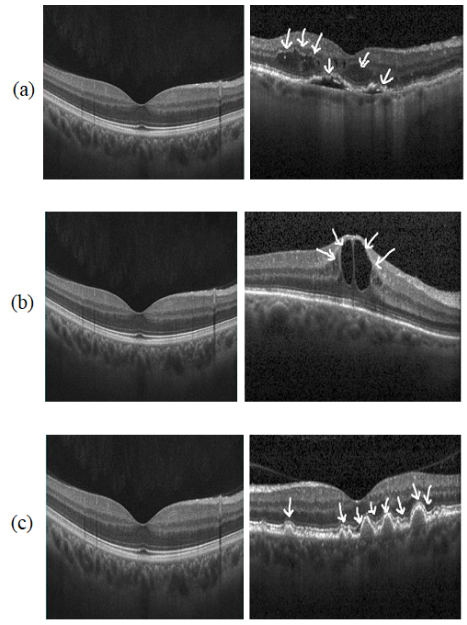

Fig. 1. Comparison Between Different eye conditions: (a) Normal vs CNV (b) Normal vs DME (c) Normal vs DRUSEN

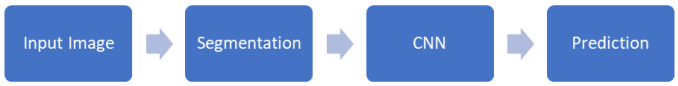

Fig. 2. Work Flow

II. Retinal OCT Scan & Diseases

A. Optical Coherence Tomography

Optical Coherence Tomography (OCT) is an advanced imaging technique which uses coherent light sources to capture images of the optical scattering media up to micrometer resolutions. The process is comparable to ultrasound. Light sources are used instead of sound waves for OCT scans. Optical Coherence Tomography has vast applications in medical imaging. With Retinal OCT scans, the detailed images are revolutionizing early detection and treatment of eye conditions such as -

- Choroidal NeoVascularization (CNV)
- Diabetic Macular Edema (DME); and
- Drusen etc.

B. Diagnosed Eye Diseases

A normal eye is free from any sort of retinal fluid and swelling/edema. Whereas, CNV is the creation of new blood vessels in the choroid layer of the eye. It is a common cause of 'wet' macular degeneration. Comparing with a normal eye, the eye with CNV has neovascular membrane and subretinal fluid.

DME (Diabetic Macular Edema) is the condition when fluid and protein deposits collect on or under the macula of the eye and causes it to thicken and swell (edema). The swelling may distort a person's central vision. The retinal-thickening-associated with intra retinal fluid can be distinguished from a normal eye in a retinal OCT.

Drusen is the tiny yellow or white accumulations of extracellular material that build up between Bruch's membrane and the retinal pigment epithelium of the eye. Fig. 1 shows comparison between different eye conditions i.e. CNV, DME, Drusen with respect to a normal eye.

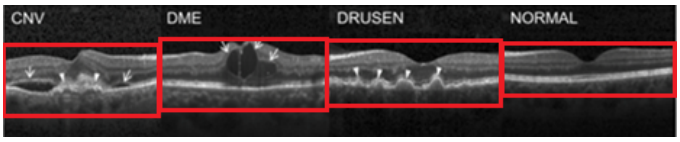

Fig. 3. Region of Interests

C. Dataset

We used the open source Optical Coherence Tomography Images for Classification [8] dataset to train our network. Dataset of the validated OCT images are described and analyzed by Kermany et al. [10]. The OCT Images are split into a training set and a testing set of independent patients. OCT Images are labeled and split into 4 classes: CNV, DME, DRUSEN, and NORMAL.

The dataset has a train set containing 83,484 images; having 37,205 images of CNV, 11,348 images of DME, 8,616 of Drusen and 26,315 images of normal eye condition. A separate test set to benchmark the performance of the network consists of 968 images, containing 242 images of each class. Since the dataset already consisted of augmented retinal OCT scans, further augmentation was not carried out.

III. Work flow

Fig. 2 shows a basic outline of our approach for the task of disease classification. First, the output of the retinal OCT scan is read in as a grayscale image. The image is primarily segmented using computer vision techniques discussed in III(A) to crop out the region of interest of the retinal OCT scan. The segmented image is then resized and run through a basic convolutional neutral network previously trained on the images of the training set. Details of the neural network is discussed in III-B. Finally, the neural network provides predictions of whether the patient has any particular retinal disease.

A. Segmentation

As we can see from Fig. 3, for each type of disease, there is a specific Region of Interest where emphasis could be given to detect the retinal abnormalities. Some basic image processing techniques were applied to segment out the Region of Interest in the OCT scans. This pre-processing (segmentation) of the images will help increase accuracy, reduce data loss and improve data processing time. Python library OpenCV [9] was used for the segmenting process.

A big chunk of the images in the dataset contained white borders on the sides. Since a Binary Thresholding operation is used to determine the Region of Interest in the scan, these white borders would interfere with the segmentation process. Therefore, the bright pixels in the bordering regions were replaced with dark ones.

For thresholding the image, we used inverted Otsus Binarization [11]. Thresholding produced a binary image with dark pixels in the retinal regions and white pixels otherwise, as shown in Fig. 4(c).

However, the thresholding operation left noise in the image which was reduced using an opening operation followed by dilation. A 3×3 unit kernel was used for both operations. 2

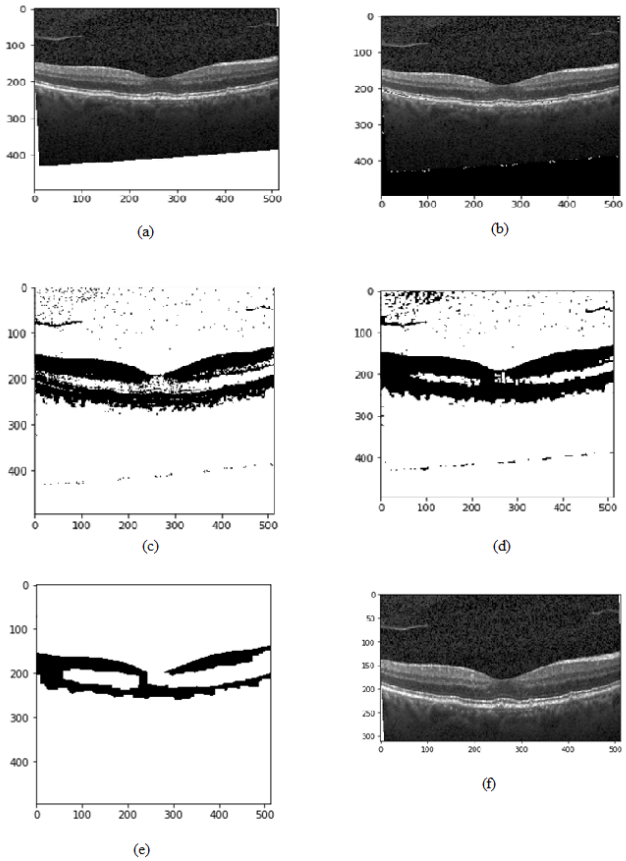

Fig. 4. (a) Sample image containing redundant white borders (b) Sample image after filling in the white borders with dark pixels (c) Sample image after Otsus Binarization (d) Sample image after opening operation (e) Sample image after further dilation (f) Cropped out Region of Interest form sample image

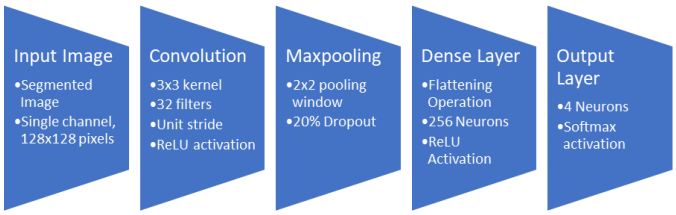

Fig. 5. Architecture of the Convolutional Neural Network

iterations of opening followed by 7 iterations of dilation provided optimum results. After the morphological transformation operations, we were left with a clean binary image, as in Fig. 4(e).

The dilated image was finally used to identify the region of interest. The minimum and maximum co-ordinates of the dark pixels along the vertical axis were used to determine the specified retinal region. However, continuous opening and dilation operation leads to shrinking of the ROI. Therefore, the ROI was expanded by 60 pixels in either direction so that unwanted cropping of the retinal region does not occur.

Finally, the region of interest was cropped out using the indices found from the previous step. A cropped-out ROI from the scan is shown in Fig. 4(f).

B. CNN Architecture

A Basic Convolutional Neural Network was used for the classification of the pre-processed Retinal OCT scan images.

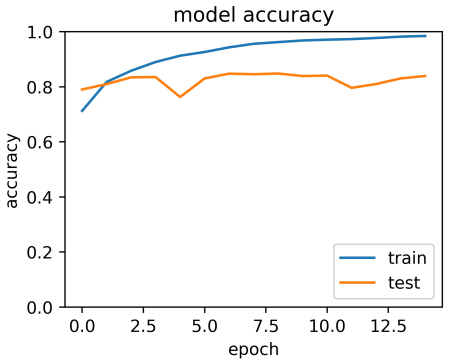

Fig. 6. Training Accuracy

Fig. 5 shows the basic structure of the Convolutional Neural Network. The pre-processed images are read as single channel gray-scale images and resized to 128×128 pixels to input to the neural network.

The first layer of the Neural Network consists of a 2D convolution operation using 32 filters having 3×3 kernels and a unit stride sliding window. Rectified-linear-unit(ReLU) is used as the activation function for the neurons in this layer. Following the convolution operation, the result is a $32 \times 126 \times 126$ tensor. A 2×2 Max-pooling operation takes place to pool the results of the convolution into a more compact $32 \times 63 \times 63$ tensor. A dropout of 20% is placed at the end of this layer to reduce over-fitting.

To connect the outputs of the max-pooled features to the fully connected layer, a flattening operation takes place to create a one-dimensional tensor of 127008 elements. The final two dense layers of the network are fully connected layers having respectively 256 and 4 neurons, the latter of which defines the output of the neural network i.e. the target class of any of the four potential diseases. The first dense layer uses Rectified-linear-unit activation while the final dense layer has the 'SoftMax' activation function for providing predictions. The final layer of the Convolutional Neural Network provides probabilities of the input image to contain each condition. The neural network provides outputs as a float point probability between 0 and 1. The class or neuron having the highest probability signifies that the patient might have the corresponding disease/abnormality.

C. Training

Fig. 6 shows the history of the accuracy of training through 15 epochs. The network converged to an accuracy of 98% on the training set and a validation accuracy as high as 85% on the holdout set within this training period.

The CNN architecture was trained using the 83,484 images of the training set. A random set of 20% of these images were kept as holdout for validating the model during the training. The network was trained using the 'Adam' optimizer [12] which is a first order gradient-based stochastic optimization procedure. The learning rate, β_1 and β_2 values were respectively set to 0.001, 0.9 and 0.999 for the training procedure. Since the number of training images were high, the model converged within 15 epochs.

TABLE I
ACCURACY METRICS

	Precision	**Recall**	**F1-Score**
Normal	0.96	0.98	0.97
CNV	0.91	1.00	0.95
DME	1.00	0.87	0.93
Drusen	0.97	0.98	0.98
Average	0.9519	0.9566	0.9562

TABLE II
CONFUSION MATRIX

		Predicted Label			
		Normal	CNV	DME	DRUSEN
True Labels	Normal	237	0	0	5
	CNV	0	241	0	1
	DME	9	20	211	2
	DRUSEN	0	5	0	237

IV. RESULTS

Accuracy metrics provided by the training of the neural network were further verified by validating classification metrics i.e. precision, recall and f1-score. Accuracy metrics on the test set for different classes are tabulated in Table I.

The results can be further elaborated by creating a confusion matrix on the predicted labels versus the true labels for the respective disease classes.

Table II shows the confusion matrix for different disease classes without normalization. The worst case of classification here is DME, which has been mis-classified 20 times as CNV and 9 times as Normal condition out of the 242 images belonging to that class, resulting in a low recall score. However, results from the other classes are quite convincing.

The class-by-class accuracy metrics of our classifier were also compared with the work of Kermany et al. [10] who use transfer learning on a 48-layer deep computationally complex Inception-v3 architecture [13] on the same dataset without any pre-processing, whereas our network only consists of a single convolution and pooling layer followed by two dense layers of neurons on the pre-processed image. Despite being computationally much simpler, our network is able to predict the retinal diseases at a nearly equivalent overall accuracy on the test set (95.66%) compared to the latter work (96.1%). A more detailed class-by-class comparision of different metrics is shown in Table III.

V. CONCLUSION

Manual classification and detection of retinal diseases from retinal OCT scans is a time-consuming process. Our proposed method is able to detect the retinal diseases with 95.66% accuracy which is comparable to that of a trained medical professional [10]. The network itself is lightweight and computationally efficient compared to the other approaches [10]. Therefore, our solution might have great practical implications and be able to aid medical professionals without expensive hardware, saving precious resources and time in the process.

TABLE III
ACCURACY COMPARISON

	Normal		Drusen		CNV		DME	
	I	II	I	II	I	II	I	II
Accuracy	98.55	98.60	98.65	97.80	97.31	97.80	96.79	98.00
Error Rate	1.45	1.40	1.35	2.20	2.69	2.20	3.21	2.00
Precision	96.34	98.40	96.73	94.40	90.60	96.80	100.00	94.80
Sensitivity	97.93	98.65	97.93	96.72	99.58	93.80	87.19	97.13
Specificity	98.76	99.46	98.90	98.14	96.55	98.92	100.00	98.28
FPR	1.24	5.40	1.10	1.86	3.45	1.08	0.00	1.72

REFERENCES

[1] Girard and Cheriet, "Artery/vein classification in fundus images using CNN and likelihood score propagation," *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 720-724.

[2] Gleryz and Ulusoy, "Retinal vessel segmentation using convolutional neural networks," *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4.

[3] Sambaturu, Srinivasan, Prabhu, Rajamani, Palanisamy, Haritz and Singh, "A novel deep learning based method for retinal lesion detection," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 33-37.

[4] Calimeri, Marzullo, Stamile and Terracina, "Optic disc detection using fine-tuned convolutional neural networks," *2016 12th International Conference on Signal-Image Technology & Internet- Based Systems (SITIS)*, pp. 69-75.

[5] Triwijoyo, Heryadi, Lukas, Ahmad, Sabarguna, Budiharto and Abdurachman, "Retina disease classification based on color fundus images using convolutional neural networks," *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, pp. 1-4.

[6] Athar, Vahadane, Joshi and Dastidar, "Weakly supervised fluid filled region localization in retinal OCT scans," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1467-1470.

[7] Gopinath and Sivaswamy, "Segmentation of retinal cysts from Optical Coherence Tomography volumes via selective enhancement," *IEEE Journal of Biomedical and Health Informatics*, pp. 1-1.

[8] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification, Mendeley Data, v2

[9] Bradski,G, "The OpenCV Library", *Dr. Dobb's Journal of Software Tools(2000)*.

[10] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K.Prasadha, Jacqueline Pei, Magdalene Y.L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A.N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A.Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, Kang Zhang, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, Volume 172, Issue 5, p. 1122-1131.e9.

[11] Rafael C. Gonzalez, *Digital Image Processing*.

[12] Kingma, Diederik and Ba, Jimmy, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations*, 2014.

[13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826.